

# A mu-differentiable Lagrange multiplier rule

Ricardo Almeida and Delfim F. M. Torres

Department of Mathematics  
University of Aveiro  
3810-193 Aveiro, Portugal  
{ricardo.almeida, delfim}@ua.pt

## ABSTRACT

*We present some properties of the gradient of a mu-differentiable function. The Method of Lagrange Multipliers for mu-differentiable functions is then exemplified.*

**Keywords:** nonstandard analysis, mu-differentiability, Method of Lagrange Multipliers.

**2000 Mathematics Subject Classification:** 26E35, 26E05 , 26B05.

## 1 Introduction

In [1] we introduce a new kind of differentiation, what we call mu-differentiability, and we prove necessary and sufficient conditions for the existence of extrema points. For the necessary background on Nonstandard Analysis and for notation, we refer the reader to [1] and references therein. Here we just recall the necessary results.

**Definition 1.1.** [1] Given an internal function  $f : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$ , we say that  $\alpha \in \mathbb{R}^n$  is a *local m-minimum* of  $f$  if

$$f(x) \gtrsim f(\alpha) \text{ for all } x \in {}^*B_r(\alpha),$$

where  $r \in \mathbb{R}$  is a positive real number.

The crucial fact is that there exists a relationship between m-minimums and minimums:

**Lemma 1.1.** [1] If  $f : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$  is mu-differentiable, then

$\alpha$  is a m-minimum of  $f$  if and only if  $\alpha$  is a minimum of  $st(f)$ .

With this lemma, and using the fact that

$$st\left(\frac{\partial f}{\partial x_i}\Big|_{\alpha}\right) = \frac{\partial st(f)}{\partial x_i}\Big|_{\alpha} \text{ for } i \in \{1, \dots, n\}, \quad (1.1)$$

it follows:

**Theorem 1.2.** [1] If  $f : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$  is a mu-differentiable function and  $\alpha$  is a m-minimum of  $f$ , then

$$\frac{\partial f}{\partial x_i}\Big|_{\alpha} \approx 0, \text{ for every } i = 1, \dots, n.$$

In this paper we develop further the theory initiated in [1], proving some properties of the gradient vector (section 2) and a Method of Lagrange Multipliers (section 3). Several examples show the analogy with the classical case.

## 2 The Gradient Vector

In the sequel  $f$  denotes an internal mu-differentiable function from  ${}^*\mathbb{R}^n$  to  ${}^*\mathbb{R}$ .

**Definition 2.1.** A *gradient vector* of  $f$  at  $x \in ns({}^*\mathbb{R}^n)$  is defined by

$$\nabla f(x) := \left( \left. \frac{\partial f}{\partial x_1} \right|_x, \dots, \left. \frac{\partial f}{\partial x_n} \right|_x \right)$$

where

$$\left. \frac{\partial f}{\partial x_i} \right|_x \approx \frac{f(x_1, \dots, x_{i-1}, x_i + \epsilon, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)}{\epsilon}$$

and  $\epsilon$  is an infinitesimal satisfying  $|\epsilon| > \delta_f$ .

*Remark 2.1.* The positive infinitesimal  $\delta_f$  that appears in Definition 2.1 is given by the m-differentiability of  $f$  (cf. [1]).

*Remark 2.2.* Observe that

$$\left. \frac{\partial f}{\partial x_i} \right|_x \approx Df_x(e_i),$$

where  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  denotes the  $i$ th canonical vector, and  $Df_x$  denotes the derivative operator of  $f$  at  $x$ .

**Theorem 2.1.** If  $x, y \in ns({}^*\mathbb{R}^n)$  and  $x \approx y$ , then

$$\left. \frac{\partial f}{\partial x_i} \right|_x \approx \left. \frac{\partial f}{\partial x_i} \right|_y, \quad i = 1, \dots, n,$$

i.e.,  $\nabla f(x) \approx \nabla f(y)$ .

*Proof.* Simply observe that  $Df_x(e_i) \approx Df_y(e_i)$ . □

**Theorem 2.2.** If  $u \in {}^*\mathbb{R}^n$  is a finite vector, then

$$\forall x \in ns({}^*\mathbb{R}^n) \quad Df_x(u) \approx \nabla f(x) \cdot u.$$

*Proof.* Since  $st(f)$  is a  $C^1$  function, it follows that for any  $v \in \mathbb{R}^n$

$$Dst(f)_{st(x)}(v) = \nabla st(f)(st(x)) \cdot v.$$

By the Transfer Principle of Nonstandard Analysis, it still holds for  $u \in {}^*\mathbb{R}^n$ . On the other hand,

1.  $Dst(f)_{st(x)}(v) = st(Df_{st(x)})(u) \approx Df_x(u),$
2.  $\nabla st(f)(st(x)) = st(\nabla f(st(x))) \approx \nabla f(x),$

which proves the desired. □

We point out that, in opposite to classical functions, if  $\nabla f(x)$  is a gradient vector of  $f$  at  $x$ , then  $\nabla f(x) + \Omega$ , where  $\Omega \in {}^*\mathbb{R}^n$  is an infinitesimal vector, is also a gradient vector at  $x$ . Conversely, if  $\nabla f(x)$  and  $\nabla^1 f(x)$  are two gradient vectors, then  $\nabla f(x) - \nabla^1 f(x) \approx 0$ .

From now on, when there is no danger of confusion, we simply write  $\nabla f$  instead of  $\nabla f(x)$ .

*Example 2.1.* Let  $f(x, y, z) = (1 + \epsilon)xy^2 - \delta z$ , with  $(x, y, z) \in {}^*\mathbb{R}^3$ , and  $\epsilon$  and  $\delta$  two infinitesimal numbers. Given an infinitesimal  $\theta$ ,

$$\begin{aligned}\frac{(1 + \epsilon)(x + \theta)y^2 - \delta z - ((1 + \epsilon)xy^2 - \delta z)}{\theta} &= (1 + \epsilon)y^2, \\ \frac{(1 + \epsilon)x(y + \theta)^2 - \delta z - ((1 + \epsilon)xy^2 - \delta z)}{\theta} &= 2(1 + \epsilon)xy + \theta(1 + \epsilon)x, \\ \frac{(1 + \epsilon)xy^2 - \delta(z + \theta) - ((1 + \epsilon)xy^2 - \delta z)}{\theta} &= -\delta,\end{aligned}$$

and we can choose

$$\frac{\partial f}{\partial x} = (1 + \epsilon)y^2, \quad \frac{\partial f}{\partial y} = 2(1 + \epsilon)xy, \quad \frac{\partial f}{\partial z} = -\delta.$$

**Theorem 2.3.** *If  $f$  and  $g$  are mu-differentiable and  $k \in \text{fin}({}^*\mathbb{R})$ , then*

$$\nabla(kf) = k\nabla f, \quad \nabla(f + g) = \nabla f + \nabla g, \quad \text{and} \quad \nabla(fg) = f\nabla g + g\nabla f.$$

*Proof.* We prove only the last equality. Fix an infinitesimal number  $\epsilon$  such that  $|\epsilon| > \delta_f$ . Then,

$$\begin{aligned}\frac{\partial(fg)}{\partial x_i} &\approx \frac{(fg)(x_1, \dots, x_{i-1}, x_i + \epsilon, x_{i+1}, \dots, x_n) - (fg)(x_1, \dots, x_n)}{\epsilon} \\ &= f(x_1, \dots, x_n) \frac{g(x_1, \dots, x_{i-1}, x_i + \epsilon, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_n)}{\epsilon} \\ &\quad + g(x_1, \dots, x_{i-1}, x_i + \epsilon, x_{i+1}, \dots, x_n) \frac{f(x_1, \dots, x_{i-1}, x_i + \epsilon, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)}{\epsilon} \\ &\approx f(x) \frac{\partial g}{\partial x_i} + g(x) \frac{\partial f}{\partial x_i}\end{aligned}$$

by the continuity of  $g$ . □

**Definition 2.2.** We say that  $x$  is a  $m$ -critical point of  $f$  if  $\nabla f(x) \approx 0$ .

The following lemma is an immediate consequence of (1.1) and Definition 2.2.

**Lemma 2.4.** *A point  $x$  is a  $m$ -critical point of  $f$  if and only if  $st(x)$  is a critical point of  $st(f)$ .*

### 3 The Method of Lagrange Multipliers

Let  $f : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$  and  $g_j : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$ ,  $j = 1, \dots, m$  ( $m \in \mathbb{N}$ ,  $m < n$ ), denote internal mu-differentiable functions. We address the problem of finding  $m$ -minimums or  $m$ -maximums of  $f$ , subject to the conditions  $g_j(x) \approx 0$ , for all  $j$ . The constraints  $g_j(x) \approx 0$ ,  $j = 1, \dots, m$ , are called *side conditions*. Lagrange solved this problem (for standard differentiable functions), introducing new variables,  $\lambda_1, \dots, \lambda_m$ , and forming the augmented function

$$F(x, \lambda_1, \dots, \lambda_m) = f(x) + \sum_{j=1}^m \lambda_j g_j(x), \quad x \in \mathbb{R}^n.$$

Roughly speaking, Lagrange proved that the problem of finding the critical points of  $f$ , satisfying the conditions  $g_j(x) = 0$ , is equivalent to find the critical points of  $F$ . We present here a method to determine critical points for internal functions with side conditions, based on the *Method of Lagrange Multipliers*. Similarly to the classical setting, define

$$F : \quad {}^*\mathbb{R}^{n+m} \quad \rightarrow \quad {}^*\mathbb{R}$$

$$(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) \mapsto f(x_1, \dots, x_n) + \sum_{j=1}^m \lambda_j g_j(x_1, \dots, x_n).$$

If we define  $g := (g_1, \dots, g_m)$  and  $\lambda := (\lambda_1, \dots, \lambda_m)$ , we can simply write

$$F(x, \lambda) = f(x) + \lambda \cdot g(x). \quad (3.1)$$

**Theorem 3.1.** *[Lagrange rule in normal form with one constraint] Let  $f : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$  and  $g : {}^*\mathbb{R}^n \rightarrow {}^*\mathbb{R}$  be two mu-differentiable functions, and  $\alpha$  a m-minimum of  $f$  such that  $g(\alpha) \approx 0$  and  $\nabla g(\alpha) \not\approx 0$ . Then, there exists a finite  $\lambda \in {}^*\mathbb{R}$  such that*

$$\nabla f(\alpha) + \lambda \nabla g(\alpha) \approx 0.$$

*Proof.* Since  $st(f)$  and  $st(g)$  are functions of class  $C^1$ ,  $\alpha$  is a minimum of  $st(f)$ ,  $st(g)(\alpha) = 0$  and  $\nabla st(g)(\alpha) \neq 0$ , it follows (see, e.g., [2, p. 148]) that

$$\exists \lambda \in \mathbb{R} \quad \nabla st(f)(\alpha) + \lambda \nabla st(g)(\alpha) = 0.$$

Hence,

$$\nabla f(\alpha) + \lambda \nabla g(\alpha) \approx 0.$$

□

*Remark 3.1.* Suppose that we are in the conditions of Theorem 3.1. Then, there exists some  $\lambda_1 \in fin({}^*\mathbb{R})$  such that

$$\nabla f(\alpha) + \lambda_1 \nabla g(\alpha) \approx 0,$$

i.e.,

$$\left. \frac{\partial f}{\partial x_i} \right|_{\alpha} + \lambda_1 \left. \frac{\partial g}{\partial x_i} \right|_{\alpha} \approx 0, \quad i = 1, \dots, n.$$

Using the notation (3.1), if  $\alpha$  is a m-minimum of  $f$  and  $g(\alpha) \approx 0$ , then

$$\begin{cases} \left. \frac{\partial F}{\partial x_i} \right|_{(\alpha, \lambda_1)} = \left. \frac{\partial f}{\partial x_i} \right|_{\alpha} + \lambda_1 \left. \frac{\partial g}{\partial x_i} \right|_{\alpha} \approx 0, & i = 1, \dots, n, \\ \left. \frac{\partial F}{\partial \lambda} \right|_{(\alpha, \lambda_1)} \approx g(\alpha) \approx 0. \end{cases} \quad (3.2)$$

Consequently, the m-minimum points are solutions of the system

$$\frac{\partial F}{\partial x_i} \approx 0, \quad i = 1, \dots, n, \quad \text{and} \quad \frac{\partial F}{\partial \lambda} \approx 0,$$

i.e.,  $\nabla F \approx 0$ .

*Example 3.1.* Let  $f(x, y, z) = xyz + \epsilon$ , with  $\epsilon \approx 0$ , and consider the constraint  $g(x, y, z) = x^2 + 2(y + \delta)^2 + 3z^2 - 1$ , with  $\delta \approx 0$ . In this case, we define

$$F(x, y, z, \lambda) := xyz + \epsilon + \lambda(x^2 + 2(y + \delta)^2 + 3z^2 - 1).$$

The system (3.2) takes the form

$$\begin{cases} yz + 2\lambda x \approx 0 \\ xz + 4\lambda(y + \delta) \approx 0 \\ xy + 6\lambda z \approx 0 \\ x^2 + 2(y + \delta)^2 + 3z^2 - 1 \approx 0. \end{cases}$$

Since

$$xyz \approx -2\lambda x^2 \approx -4\lambda y(y + \delta) \approx -6\lambda z^2,$$

if  $\lambda \not\approx 0$ , the solution is

$$x^2 \approx \frac{1}{3}, \quad y^2 \approx \frac{1}{6} \text{ and } z^2 \approx \frac{1}{9};$$

if  $\lambda \approx 0$ , then

$$\left(0, 0, \pm \frac{1}{\sqrt{3}}\right), \left(0, \pm \frac{1}{\sqrt{2}}, 0\right) \text{ and } (\pm 1, 0, 0)$$

are solutions. Observe that

$$\nabla g = (2x, 4(y + \delta), 6z) \approx (0, 0, 0) \text{ if and only if } (x, y, z) \approx (0, 0, 0).$$

One easily checks that

$$f\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{6}}, \frac{1}{3}\right) = \frac{1}{3\sqrt{18}} + \epsilon \text{ is the m-maximum and}$$

$$f\left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{6}}, \frac{1}{3}\right) = -\frac{1}{3\sqrt{18}} + \epsilon \text{ is the m-minimum}$$

of  $f$  subject to the constraint  $g$ .

We now prove a more general Lagrange rule, admitting possibility of abnormal critical points ( $\mu = 0$ ) and multiple constraints.

**Theorem 3.2 (Lagrange rule).** *Let  $f, g_1, \dots, g_m$  be  $m$ -differentiable functions on  ${}^*\mathbb{R}^n$ . Let  $\alpha$  be a  $m$ -minimum of  $f$  satisfying*

$$g_1(\alpha) \approx \dots \approx g_m(\alpha) \approx 0.$$

*Then, there exist finite hyper-reals  $\mu, \lambda_1, \dots, \lambda_m \in {}^*\mathbb{R}$ , not all infinitesimals, such that*

$$\mu \nabla f(\alpha) + \lambda_1 \nabla g_1(\alpha) + \dots + \lambda_m \nabla g_m(\alpha) \approx 0.$$

*Remark 3.2.* Defining  $F(x, \mu, \lambda) := \mu f(x) + \lambda \cdot g(x)$ , the necessary optimality condition given by Theorem 3.2 can be written as  $\partial F / \partial x \approx \partial F / \partial \lambda \approx 0$ .

*Proof.* First observe that  $st(f), st(g_1), \dots, st(g_m)$  are all functions of class  $C^1$ ,  $\nabla st(f)(\alpha) = st(\nabla f)(\alpha)$  and  $\nabla st(g_j)(\alpha) = st(\nabla g_j)(\alpha)$ , for  $j = 1, \dots, m$ . Furthermore, since  $\alpha$  is a minimum of  $st(f)$  and

$$st(g_1)(\alpha) = \dots = st(g_m)(\alpha) = 0,$$

there exist reals  $\mu, \lambda_1, \dots, \lambda_m$ , not all zero, such that

$$\mu \nabla st(f)(\alpha) + \lambda_1 \nabla st(g_1)(\alpha) + \dots + \lambda_m \nabla st(g_m)(\alpha) = 0$$

(see, e.g., [2, p. 148]). Consequently,

$$\mu st(\nabla f)(\alpha) + \lambda_1 st(\nabla g_1)(\alpha) + \dots + \lambda_m st(\nabla g_m)(\alpha) = 0. \quad (3.3)$$

On the other hand, we have

$$\mu st(\nabla f)(\alpha) = \mu st(\nabla f(\alpha)) \approx \mu \nabla f(\alpha).$$

Analogously, for each  $j = 1, \dots, m$ ,

$$\lambda_j st(\nabla g_j)(\alpha) \approx \lambda_j \nabla g_j(\alpha).$$

Substituting on equation (3.3) the previous relations, one proves the desired result.  $\square$

*Example 3.2.* Let  $f(x, y, z) = z^2/2 - (x + \epsilon)y$ , with  $\epsilon \approx 0$ , be the function to be extremized, and  $g_1(x, y, z) = x^2 + y - 1$  and  $g_2(x, y, z) = x + z - 1 + \delta$ , with  $\delta \approx 0$ , be the constraints. Then, the augmented function is

$$F(x, y, z, \mu, \lambda_1, \lambda_2) = \mu [z^2/2 - (x + \epsilon)y] + \lambda_1(x^2 + y - 1) + \lambda_2(x + z - 1 + \delta).$$

To find the local extrema of  $f$ , subject to the conditions  $g_1 \approx 0$  and  $g_2 \approx 0$ , we form the system

$$\begin{cases} -\mu y + 2\lambda_1 x + \lambda_2 \approx 0 \\ -\mu(x + \epsilon) + \lambda_1 \approx 0 \\ \mu z + \lambda_2 \approx 0 \\ x^2 + y - 1 \approx 0 \\ x + z - 1 + \delta \approx 0 \end{cases} \quad (3.4)$$

of necessary optimality conditions. Assume  $\mu \approx 0$  (abnormal case). Then, the first two equations in (3.4) imply immediately that  $\lambda_1 \approx \lambda_2 \approx 0$ . This is not a possibility by Theorem 3.2. We conclude that  $\mu \not\approx 0$ . The solutions of (3.4) are then infinitely close to the vectors

$$(-1, 0, 2) \quad \text{and} \quad (2/3, 5/9, 1/3).$$

Hence, if  $f$  has any m-extrema under the given constraints, then they must occur at either  $(-1, 0, 2)$  or  $(2/3, 5/9, 1/3)$ .

## Acknowledgment

This work was supported by *Centre for Research on Optimization and Control* (CEOC) from the “Fundação para a Ciência e a Tecnologia” (FCT), cofinanced by the European Community Fund FEDER/POCI 2010.

## References

- [1] R. Almeida and D. F. M. Torres. Relaxed optimality conditions for mu-differentiable functions. *Int. J. Appl. Math. Stat.* (submitted).
- [2] W. Cheney. *Analysis for Applied Mathematics*. GTM. Springer-Verlag, New York, 2001.